6th Asian-Pacific Symposium on Structural Reliability and its Applications (APSSRA2016)
28-30 May 2016, Shanghai, China
H.W. Huang, J. Li, J. Zhang & J.B. Chen (editors)

Spectral Moments Estimates Uncertainty Quantification under Incomplete Data

Yuanjin Zhang[1*], Liam Comerford[1], Ioannis A. Kougioumtzoglou[2], Edoardo Patelli[1] and Michael Beer[3]

[1*] *Institute for Risk & Uncertainty, University of Liverpool, Liverpool, UK*
ylzhyj@liverpool.ac.uk; phdliam@liverpool.ac.uk; epatelli@liverpool.ac.uk; mbeer@liverpool.ac.uk
[2] *Dep. of Civil Engineering and Engineering Mechanics, Columbia University, New York, USA,*
ikougioum@columbia.edu
[3] *Faculty of Civil Engineering, Leibniz University Hannover, Hannover, Germany,*
beer@bauinf.uni-hannover.de

Abstract: In general, classical stochastic process power spectrum estimation techniques with missing data typically provide a deterministic estimate for the power spectrum value. Thus, no information is provided concerning the uncertainty in the estimates. The significance of the derived PDF relates to cases where incomplete process realizations are available for power spectrum estimation applications. Examples presented herein demonstrate the large extent to which any given single deterministic estimate, even for small amounts of missing data, may be unrepresentative of the real spectrum. In this paper, the issue of quantifying the uncertainty in stochastic process power spectrum estimates based on realizations with missing data is addressed. Relying on relatively relaxed assumptions for the missing data, utilizing fundamental concepts from probability theory, and resorting to a Fourier based representation of stationary stochastic processes, a closed-form expression for the probability density function of the power spectrum value corresponding to a specific frequency is derived. The given approach is then extended for assessing the uncertainty in spectral moment estimates based on a Kriging extrapolation scheme. Finally, the distribution function of the power spectrum value corresponding to a specific frequency is determined while considering correlations between the missing points. In this manner, the propagation of the uncertainty from the available realizations to the spectral moment estimates can be readily monitored and quantified.

Keywords: spectral moments, missing data, kriging, spectral estimation

1. Introduction

In areas such as stochastic dynamics, stochastic processes are described by statistical properties such as the power spectrum. There are numerous approaches to process spectrum estimation in the literature. For example, a Fourier basis may be used in the spectral estimation of stationary processes (Newland, 1993); similar to the stationary case, nonstationary processes can be estimated by utilizing a wavelet basis (Spanos P. D. and Failla G., 2004) or a Gaussian chirplet basis (Politis, et al., 2007). These spectral estimation methods often require a large number of full set samples to achieve an adequate predefined degree of accuracy.

However, the presence of missing data in measurements is frequently unavoidable; in fact, missing data are possible in almost any situation where the data are collected and stored. In seismic or wind measurement processes, data missing is often a big issue to tackle. Reasons for which include equipment failure or restricted use of equipment, as well as data corruption and cost / bandwidth limitations to name just a few. Thus, in such cases standard spectral analysis techniques cannot be used for spectral estimation in a straightforward manner, especially for those methods based on the standard Fourier transformation, harmonic wavelet transformation and chirplet transformation, all of which are based on the manipulation of full set data. In this regard, many efforts have been made to mitigate the effect of missing data in spectral estimation according to the different degrees of missing data amount and types of prior knowledge. Comerford et al. (2015a.) developed a compressive sensing technique for estimating power spectra based on stochastic processes compatible realizations with missing data. This is especially suitable for treating multiple records of a single process, where a reweighting procedure could be introduced to improve the result to a large degree. Note that the main assumption relates to the records having a sparse representation of a known basis; see also Fahlman and Ulrych (1982), and Comerford et al. (2015) for other alternative techniques for dealing with missing data.

Although all the methods mentioned above could, depending on the setting, potentially provide a relatively accurate estimation of power spectrum, they will also propagate inaccuracies from missing data predictions in the time domain through to the final spectral estimates. Most of the aforementioned methods estimate the power spectrum by reconstructing missing parts of the data, and based on these reconstructed full data, standard spectral analysis methods are involved. The level of accuracy depends on the confidence in the specific method as the uncertainty of the missing data is rarely taken into consideration. Hence, a reliable method of quantifying the impact of missing data on the uncertainty in a given spectral estimate could be a quite useful tool.

To quantify the uncertainty of spectral estimation subject to missing data, a probability density function (PDF) for representing the missing data is considered (Comerford et al. 2015b). As a result, the power spectrum estimate is also determined in a probabilistic manner, i.e., the approximate PDF of each individual spectral point in the frequency domain can be calculated instead of a deterministic value. Comerford et al. (2015b). demonstrates a method for determining the "probabilistic" power spectrum with the assumption that all the variables in the time domain satisfy the same independent Gaussian distribution. However, this method does not consider the correlation between the missing points and for larger numbers of missing data, is largely unrepresentative of a signal with harmonic properties. By virtue of the central limit theorem (Billingsley, 2008), it is still reasonable to consider the missing points as multi-variable Gaussian variants, especially in the environmental excitations such as earthquake, wind, sea waves and structure responses subject to these effects. In this regard, the method in (Comerford et al. 2015b) can be extended to consider the correlation between the missing data. Indeed, an extended methodology is developed herein to not only evaluate the PDF of spectral points but also quantify the uncertainty of the spectral moments. The significance of the latter feature is important in number of applications such as the development of reliability assessment methodologies where the knowledge of spectral moments is an integral part of the approaches (e.g. first-passage problem; see Vanmarcke 1975). Unlike adopting a computationally demanding Monte Carlo simulation approach, the herein proposed method provides with explicit closed form expressions for the spectral estimate and the spectral moment PDFs. Although the exact correlation between missing points is very hard to obtain in practice, a course result could be acquired by various methods. In this paper, the correlation between missing points is estimated by the existing data via a Kriging model (Krige, 1951).

2. Mathematical formulation

2.1 Power spectrum estimations PDFs for stationary stochastic process

Consider a zero mean stationary process expressed in the general form (Cramér and Leadbetter, 1967; Priestley, 1982)

$$f(t) = \int_{-\infty}^{+\infty} A(\omega) e^{iwt} \, dZ(\omega) \tag{1}$$

where $A(\omega)$ is a deterministic function and $dZ(\omega)$ is a zero mean orthonormal increment stochastic process. The two-sided power spectrum $S_f(\omega)$ of process $f(t)$ is then defined as

$$S_f(\omega) = |A(\omega)|^2 \tag{2}$$

In general, stochastic process realizations compatible with the given spectrum could be generated by a spectral representation approach (Shinozuka and Deodatis, 1991) as follows,

$$f(t) = 2 \sum_{n=0}^{N-1} \sqrt{S_f(\omega)\Delta\omega} \cos(\omega_n t + \phi_n) \tag{3}$$

where ϕ_n is the independent random phase angle distributed uniformly over the interval $[0,2\pi]$. Note that the realizations generated via Eq. (3) are also ergodic, and thus, the spectrum

$S_f(\omega)$ can be estimated by computing the temporal mean value of the square of the discrete Fourier transform (DFT) of the available record in the form,

$$S_f(\omega_k) = \lim_{N\to\infty} \frac{T}{2\pi N^2} \left|\sum_{n=0}^{N-1} x_n e^{-2\pi ikn/N}\right|^2 \tag{4}$$

where N is the number of data points, t and k are the indices of the time and frequency respectively, and T is the time duration. In practice, only discrete data with finite length are available to measure, thus, the condition $N \to \infty$ is removed with the assumption that length is relatively long enough to give an accurate estimation of the spectrum. Following the notation of (Comerford et al. 2015b), the data points are divided into 2 parts: the known points x_α and missing points x_β, where α and β are indices of the known and unknown points respectively. Further, utilizing Eq.(4), the spectrum estimate can be written in the matrix form

$$S_f(\omega_k) = (c_1 + a' X_\beta)^2 + (c_2 + b' X_\beta)^2 \tag{5}$$

$$c_1 = \sqrt{\frac{T}{2\pi N^2}} \sum_\alpha x_\alpha \cos\left(\frac{2\pi k\alpha}{N}\right) \tag{6}$$

$$c_2 = \sqrt{\frac{T}{2\pi N^2}} \sum_\alpha x_\alpha \sin\left(\frac{2\pi k\alpha}{N}\right) \tag{7}$$

$$a = \sqrt{\frac{T}{2\pi N^2}} \left[\cos\left(\frac{2\pi k\beta_1}{N}\right), \cos\left(\frac{2\pi k\beta_2}{N}\right), \dots, \cos\left(\frac{2\pi k\beta_u}{N}\right)\right]' \tag{8}$$

$$b = \sqrt{\frac{T}{2\pi N^2}} \left[\sin\left(\frac{2\pi k\beta_1}{N}\right), \sin\left(\frac{2\pi k\beta_2}{N}\right), \dots, \sin\left(\frac{2\pi k\beta_u}{N}\right)\right]' \tag{9}$$

$$X_\beta = \left(x_{\beta 1}, x_{\beta 2}, \dots x_{\beta u}\right)' \tag{10}$$

where u is the number of missing points.

By virtue of the central limit theorem (Billingsley, 2008), it is still reasonable to consider the missing points as multi-variable Gaussian variants, especially in environmental excitations such as earthquake, wind, sea waves and structure responses subject to these effects. It is assumed next that the mean μ and correlation matrix Σ of missing data are obtained by any available methodology, such as a Kriging model; this is discussed in section 2.3. Overall, the random variables are satisfying a Gaussian distribution, i.e. $X_\beta \sim N(\mu, \Sigma)$. Treating X_β as multiple Gaussian random variables (Papoulis and Pillai, 2002), Eq. (5) could be rearranged as a function of two variables

$$\begin{aligned} S_f(\omega_k) &= \left(c_1 + a' X_\beta\right)^2 + \left(c_2 + b' X_\beta\right)^2 \\ &= {X_1}^2 + {X_2}^2 \end{aligned} \tag{11}$$

It is easily confirmed that $X_1 = c_1 + a' X_\beta \sim N(c_1 + a'\mu, a'\Sigma a)$ and $X_2 = c_2 + b' X_\beta \sim N(c_2 + a'\mu, b'\Sigma b)$ form 2-dimensional joint Gaussian variables (Papoulis and Pillai, 2002). Because both X_1 and X_2 are represented by the same set of random variables X_β, it is obvious that they exhibit some degree of correlation. In this regard, the correlation matrix $C_{X_1X_2}$ of joint Gaussian variables X_1 and X_2 can be described as

$$C_{X_1X_2} = \begin{pmatrix} a'\Sigma a & \sum_i \sum_j a_i b_j (\Sigma_{ij} + \mu_i\mu_j) - b'\mu a'\mu \\ \sum_i \sum_j a_i b_j (\Sigma_{ij} + \mu_i\mu_j) - b'\mu a'\mu & b'\Sigma b \end{pmatrix} \tag{12}$$

whereas the mean vector of joint Gaussian variables X_1 and X_2 can be described as

$$\mu_{X_1X_2} = (c_1 + \mu, c_2 + \mu)' \tag{13}$$

Further, to determine the PDF of variable $S_f(\omega_k)$ in Eq. (11), the celebrated input-output PDF relationship (Papoulis and Pillai, 2002), is applied. The cumulative distribution function (CDF) of $S_f(\omega_k)$ is defined as

$$\begin{aligned} F(S_f) &= P(S_f \le s) = P[(X_1, X_2) \in D_z] \\ &= \iint_{(x_1,x_2)\in D_z} f_{X_1,X_2}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \end{aligned} \tag{14}$$

and the PDF of $S_f(\omega_k)$ is defined as

$$f_S(s) = \frac{\mathrm{d}F(S_f)}{\mathrm{d}s} \tag{15}$$

where $f_{X_1,X_2}(x_1, x_2)$ is the distribution function of variables X_1 and X_2. From Eqs. (14)—(15), an analytical expression for the PDF for the power spectrum estimate at a given frequency ω_k is given as

$$p_{S_f(\omega_k)}(s) = \frac{\mathrm{d}}{\mathrm{d}s}\left\{\iint_{x_1{}^2+x_2{}^2\leq s}\frac{1}{2\pi\sqrt{|C_{X_1X_2}|}}\exp\left[-\frac{1}{2}(X - \mu_{X_1X_2})' C_{X_1X_2}{}^{-1}\left(X-\mu_{X_1X_2}\right)\right]\mathrm{d}x_1\mathrm{d}x_2\right\} \quad (16)$$

Here, a method of determining the uncertainty of individual spectral values subject to Gaussian distributed missing data has been demonstrated as an explicit expression Eq. (16). Compared with the approach in (Comerford et al. 2015b) which makes the assumption that the missing data in a realization are independent identically distributed Gaussian variables, the assumption of missing data herein is more relaxed (and perhaps more realistic in certain cases) in the sense that missing data are only assumed as Gaussian. In this manner, the approach in (Comerford et al. 2015b) is generalized, and the correlation between the missing data is also taken into consideration in estimating the power spectrum PDF. In cases where there are more than one samples subject to missing data being measured, the spectrum estimation function Eq. (4) needs to be recast into the form

$$S_f(\omega_k) = E\left(\frac{T}{2\pi N^2}\left|\sum_{n=0}^{N-1} x_n e^{-2\pi ikn/N}\right|^2\right) \quad (17)$$

where the expectation operator is understood in an ensemble average form.

2.2 Estimating the PDF spectral moments for stationary stochastic processes

For stationary random processes, the spectral moments are defined as

$$\lambda_i = \int_{-\infty}^{+\infty} \omega^i S(\omega)d\omega \quad (18)$$

where $S(\omega)$ is the two-sided spectral density function (Vanmarcke, 1975). It is well known that the zero spectral moment λ_0 is equal to the mean square of the process $\sigma_X{}^2$, and the second spectral moment λ_2 is the mean square derivative $\sigma_{\dot{X}}{}^2$. Specifically, if process X is the structural system response, the moments λ_0 and λ_2 are the displacement and velocity variances, respectively. Spectral moments can be useful in a number of applications such as to approximately determine the distribution of first passage time for normal stationary random processes (Vanmarcke, 1975; Cramer and Leadbetter, 1967; Rice, 1945).
In general, Eq. (18) can be recast into a discrete form in the frequency domain

$$\lambda_i = \sum_n \omega_n{}^i S(\omega_n)\Delta\omega \quad (19)$$

Note that, first, spectral points satisfy different distributions according to the result in the last section, making it very difficult to reduce the number of random variables. Second, these spectral points exhibit correlation, as they are established by the same set of random variables.

Next, the characteristic function is utilized to address these challenges. By definition, the characteristic function of a random variable has a close relationship with its Fourier transformation. For a random variable, its characteristic function is defined as (Papoulis and Pillai, 2002)

$$\Phi_X(\omega) = E\left[e^{i\omega x}\right] = \int_{-\infty}^{+\infty} f_X(x)e^{i\omega x}\,\mathrm{d}x \quad (20)$$

where $f_X(x)$ is the probability density function of X. Clearly, the characteristic function and probability density function form a Fourier transform pair. Further, the spectral moments function Eq. (19) could be viewed as a quadratic transformation of missing points X_β. This is briefly described as following.

Herein, for simplicity, correlated variables $X_\beta \sim N(\mu, \Sigma)$ could be replaced by a new set of independent standard Gaussian variables $X_g \sim N_u(0, I)$

$$X_\beta = \mu + AX_g \quad (21)$$

$$\Sigma = AA' \quad (22)$$

Further, Eq. (5) could be recast into the matrix form

$$S_f(\omega_k) = (c_1 + a' X_\beta)^2 + (c_2 + b' X_\beta)^2$$

$$= (c_1 + a'\mu + a'X_g)^2 + (c_2 + b'\mu + b'X_g)^2 = X'_{gn} B X_{gn} \quad (23)$$

where

$$X_{gn} = (X'_g, 1)' = (x_{g1}, x_{g2}, \dots x_{gu}, 1)' \quad (24)$$

$$B_{ij} = \begin{cases} a_i a_j + b_i b_j \ , \ i,j \le u \\ (c_1 + a'\mu)a_i + (c_2 + b'\mu)b_i \ , \ j = u+1 \\ (c_1 + a'\mu)a_j + (c_2 + b'\mu)b_j \ , \ i = u+1 \\ (c_1 + a'\mu)^2 + (c_2 + b'\mu)^2 \ , \ i = j = u+1 \end{cases} \quad (25)$$

Combining Eqs. (19) and (20), spectral moments also take a matrix form, i.e.

$$\lambda_i = X_{gn}'(\sum_n {\omega_n}^i \Delta\omega\, B_n) X_{gn} \quad (26)$$

According to probability theory, it is straightforward to determine the characteristic function of the spectral moments; that is,

$$\Phi_{\lambda_i}(\omega) = E[e^{i\omega\lambda_i}] = \int_{-\infty}^{+\infty} (2\pi)^{-\frac{u}{2}} \exp\left\{-\frac{1}{2}\left[X'_g X_g - i\omega X'_{gn}\left(\sum_n {\omega_n}^i \Delta\omega\, B_n\right) X_{gn}\right]\right\} \mathrm{d}X_g \quad (27)$$

Computing the integral in Eq. (27) numerically is computationally intensive. However, considering the form of this integral, it is seen that it still contains a quadratic polynomial with multiple variables. Thus, to evaluate the integral,

(1) Let

$$Y = \frac{1}{2}\left(X'_g X_g - i\omega X'_{gn}\left(\sum_n {\omega_n}^i \Delta\omega\, B_n\right) X_{gn}\right) \quad (28)$$

Eq. (28) can be divided into two parts. One contains the second order terms, i.e. $Y_1 = \sum_{i,j} x_i x_j$, while the other contains the first order terms plus a constant term, i.e. $Y_2 = \sum_i x_i + c_{cons}$. Thus, Eq. (27) is rewritten into the form

$$\Phi_{\lambda_i}(\omega) = E[e^{i\omega\lambda_i}] = \int_{-\infty}^{+\infty} (2\pi)^{-\frac{u}{2}} \exp(-Y_1 - Y_2)\, \mathrm{d}X_g \quad (29)$$

(2) Similar to Eq. (23), Y_1 also has a matrix form $Y_1 = X'_g B_{Y_1} X_g$ and satisfies the condition

$$B_{Y_1} = A'_{Y_1} A_{Y_1} \quad (30)$$

where A_{Y_1} is an upper triangular matrix, and A'_{Y_1} means the transpose of A_{Y_1}. The factorization in Eq. (30) is very similar to a Cholesky factorization with the difference that the diagonal elements in B_{Y_1} are still complex values.

3) Finally, Eq. (28) can be written in the form,

$$Y = (A_Y X_{gn})'(A_Y X_{gn}) + c_Y \quad (31)$$

where $A_Y = (A_{Y_1}, a_{u\times1})$ and $a_{u\times1}$ are the coefficients corresponding to the first order terms $\sum_i x_i$ in Y_2, and c_Y is the constrant without any term of x_i.

4) Finally, the integral in Eq. (29) is calculated using Eq. (31), i.e.

$$\Phi_{\lambda_i}(\omega_k) = E[e^{i\omega_k\lambda_i}] = \sqrt{\det(B_{Y_1})}\exp(-c_Y) \quad (32)$$

And the PDF of spectral moments are estimated via the inverse Fourier transform of Eq. (32), i.e.

$$p_{\lambda_i}(s) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Phi_{\lambda_i}(\omega) e^{i\omega s}\, d\omega \quad (33)$$

The proposed method is not only suitable for the stationary process, but also for the nonstationary process provided the correlation of missing points is known.

2.3 Kriging model

The method described in the previous section considers the correlation between the missing data. There are many ways to obtain a course correlation in practice. Herein, the Kriging model (Zhang, 2005) is used to this aim.

Let $f(t)$ be the sample of a stationary stochastic process with a power spectrum S_f. Given the n known points $(t_i), i = 1,2,\dots,n$, we can obtain an estimate of $f(t_j)$ at the missing point t_j. The Kriging estimator is given by weighted linear combinations of the available known points (Zhang, 2005)

$$f(t_j) = \sum_{i=1}^{n} \lambda_i f(t_j) \quad (34)$$

where λ_i is the weight of each sample. It is shown that Eq. (34) satisfies the unbiased condition if $\sum_{i=1}^{n} \lambda_i = 1$. Under this condition, the variance of the estimate error becomes

$$V = Var[f^*(t_j) - f(t_j)] = 2\sum_{i=1}^{n} \lambda_i \gamma(t_i - t_j)$$

$$-\sum_{i=1}^{n}\sum_{k=1}^{n}\lambda_i\lambda_k\gamma(t_i-t_k) \tag{35}$$

where $\gamma(h)$ denotes the variogram. To minimize the estimate variance V, the Lagrange multiplier κ could isused to obtain the equations

$$\begin{cases}\sum_{i=1}^{n}\lambda_i\gamma(t_i-t_k)+\kappa=\gamma(t_i-t_j)\\ \sum_{i=1}^{n}\lambda_i=1\end{cases} \tag{36}$$

where $\gamma(t_i-t_k)$ is the value of variogram between the points $f(t_i)$ and $f(t_k)$. By solving Eq. (36), the weight λ_i and Lagrange multiplier κ could be calculated, then an estimate of the missing point is given by Eq. (34).

Regarding the variogram, it defines the relationship between variability and the lag distance. In general, the variogram of samples, which is also called experimental variogram or empirical variogram (Isaaks and Srivastava, 1989) can be calculated as

$$\gamma_e(h)=\frac{1}{2N_h}\sum_{i=1}^{N_h}[f(t_i+h)-f(t_i)]^2 \tag{37}$$

where N_h is the number of pairs separated by vector h, and h is the lag distance.

In theory, for stationary stochastic processes, Eq. (37) is approaching the ideal result only when there are infinite samples or the single time history length tends toward infinity. However, in practice, only limited numbers of samples are be available, hence only the course trend of the variogram is obtainable instead of the exact value. In this regard, the ideal variogram model (Webster and Oliver, 2007; Pyrcz and Deutsch, 2003) is introduced to approximately fit this trend.

In engineering applications, the variogram of many environmental excitations such as earthquakes, winds and sea waves can fluctuate more or less periodically rather than increasing monotonically, which is called the hole effect (Webster and Oliver, 2007). Herein, a non-monotonically increasing variogram model is used to describe the fluctuation of the variogram (Pyrcz and Deutsch, 2003), i.e.

$$\gamma(h)=c\left[1-e^{-\frac{3h}{d}}\cos\left(\frac{h}{a}\pi\right)\right] \tag{38}$$

where a, c, d are parameters to be determined via a least squares criterion as

$$\min_{a,c,d}|\gamma(h)-\gamma_e(h)|_2 \tag{39}$$

Thus, the correlation between the missing points is obtained by

$$\Sigma(h)=\gamma(+\infty)-\gamma(h) \tag{40}$$

Through the Kriging model, the mean values and correlation of missing data can both be obtained. Note that both the mean and variance are approximate values due to the missing data. With the mean and covariance of missing data, the approximate distribution of missing points can be obtained. Based on these estimations, the spectrum range can be estimated by the methods described in the previous sections.

3. Numerical example

For the stationary case, process records are generated compatible with the Kanai-Tajimi kind earthquake engineering power spectrum of the form

$$S(\omega)=\frac{1}{1.426}\frac{1+a\omega^2}{(\omega_g^2-\omega^2)+4\zeta_g^2\omega_g^2\omega^2} \tag{41}$$

where the natural frequency $\omega_g=12\text{rad/s}$, damping ratio $\zeta_g=0.6$ and $a=20$. The missing data are uniformly randomly distributed. To compare with the method described in (Comerford et al. 2015b), a factor $\frac{1}{1.426}$ is introduced to make the standard variance equal to 1. Fig. 1 shows the spectral ranges with 10% data missing by the method presented herein. Fig. 2 shows the result of the methodology developed in (Comerford et al. 2015b) where correlation of missing data is not taken into consideration and the missing points are independent identical Gaussian random variables, i.e. $X_\beta\sim N(0,I)$. Compared with Fig. 2, the method presented here provides a smaller range, and the mean spectrum fits the original spectrum better. Figs. 3—4 show the PDFs at different frequencies at 18, and 60 rad/s with 10% missing data replaced by correlated Gaussian random variables, compared with independent identically distributed Gaussian random variables. The vertical line shows the spectral value without missing data. Figs. 5—6 show the spectral moments λ_0 and λ_2

respectively, compared with pertinent Monte Carlo simulations.

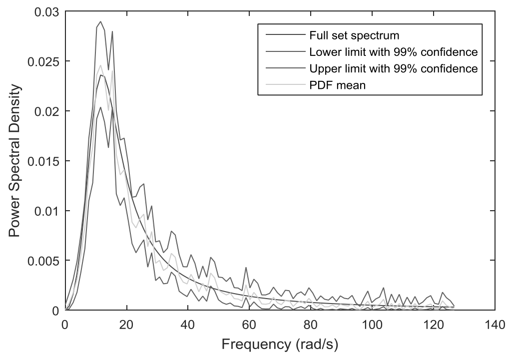

Figure 1. Power spectral probability densities with 10% missing data replaced by correlated Gaussian random variables

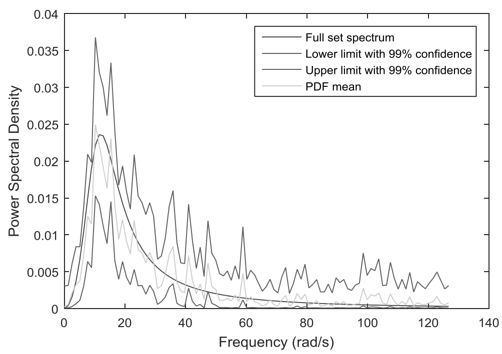

Figure 2. Power spectral probability densities with 10% missing data replaced by independent identically distributed Gaussian random variables

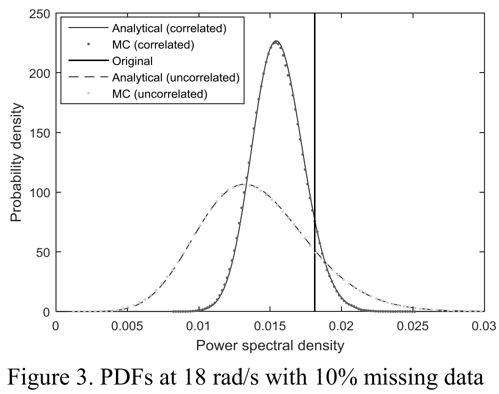

Figure 3. PDFs at 18 rad/s with 10% missing data replaced by correlated Gaussian random variables, compared with independent identically distributed Gaussian random variables. The vertical line shows the spectral value without missing data.

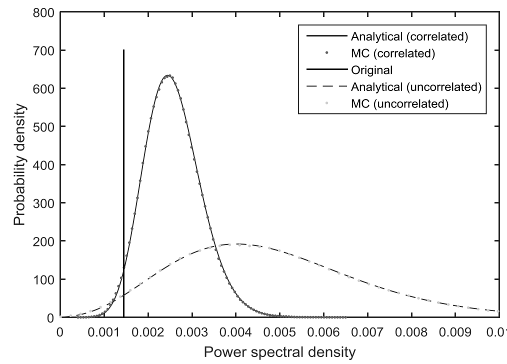

Figure 4. PDFs at 60 rad/s with 10% missing data replaced by correlated Gaussian random variables, compared with independent identically distributed Gaussian random variables. The vertical line shows the spectral value without missing data.

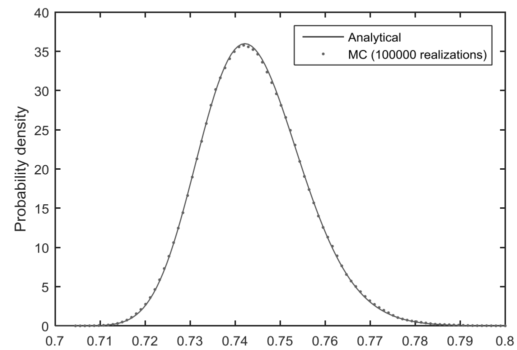

Figure 5. PDF of λ_0 with 10% missing data

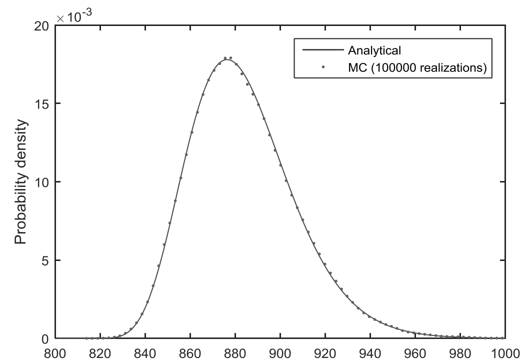

Figure 6. PDF of λ_2 with 10% missing data

4. Conclusion

In this paper, an analytical approach for quantifying the uncertainty in stochastic process power spectrum estimates based on samples with missing data has been developed. Specifically, the correlations between the missing data are considered by employing a Kriging model, while a closed form expression has been derived for the spectral estimate PDF at each frequency. The results have been compared with pertinent Monte Carlo simulations.

References

Newland. D. E. 1993. An introduction to random vibrations, spectral and wavelet analysis, *Dover Publications*.

Spanos P. D. and Failla G. 2004. Evolutionary spectra estimation using wavelets, *J. Eng. Mech*, 130: 952-960.

Politis, N. P., Giaralis, A. & Spanos, P. D. 2007. Joint time-frequency representation of simulated earthquake accelerograms via the adaptive chirplet transform, *Proc. Computational stochastic Mechanic*s, 549-557.

Comerford, Liam, Ioannis A. Kougioumtzoglou, and Michael Beer. 2015.Compressive sensing based stochastic process power spectrum estimation subject to missing data. *Probabilistic Engineering Mechanics* .

Fahlman G. and Ulrych T. 1982. A new method for estimation the power spectrum of gapped data, Monthly Notices of the Royal Astronomical Society, 199(1): 53-65.

Comerford L., Kougioumtzoglou, I.A, Beer, M. 2015a. An artificial neural network approach for stochastic process power spectrum estimation subject to missing data, *Structural Safety*, Part B, 52(0): 150-160.

Lomb, N. 1976. Least-Squares Frequency-Analysis of Unequally Spaced Data, *Astrophysics and Space Science*, 39(2) pp. 447-462.

Hays, J. D., Imbrie, J., and Shackleton, N. J. 1976. Variations in the Earth's orbit: pacemaker of the ice ages. *American Association for the Advancement of Science*.

Comerford, L.A., Kougioumtzoglou, I.A., Beer, M. 2015b. On quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data – International Journal of Sustainable Materials and Structural Systems (IJSMSS). (currently under review)

Billingsley, P. 2008. Probability and measure. John Wiley & Sons.

Vanmarcke E. H. 1975. On the distribution of the first passage time for normal stationary random process, *Journal of Applied Mechanic*s, 215-220.

Cramér, H., and Leadbetter, R. 1967. Stationary and related stochastic processes: sample function properties and their applications, Wiley.

Priestley, B. 1982. Spectral analysis and time series. Academic Press.

Shinozuka M. and Deodatis G. 1991. Simulation of stochastic processes by spectral representation, *Appl. Mech. Rev,* 44(4): 192-203.

Bracewell R. N. 1986. The Fourier transform and its applications, McGraw-Hill.

Papoulis A., Pillai S. U. 2002. Probability, Random Variables and Stochastic Processes, McGraw-Hill.

Cramer H. and Leadbetter M. R. 1967. Stationary and related stochastic processes, Wiley.

Rice S. O. 1945. Mathematical analysis of random noise, *Bell System Technical Journal,* 24-46.

Vanmarke E. H.1970. Parameters of the spectral density function: their significance in the time and frequency domains, *MIT Research Report*, R70-85.

Vanmarke E. H. 1972. Properties of spectral moments with applications to random vibration, *Journal of Engineering Mechanics Division*, 98: 425.

Krige D. G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. of the Chem., Metal. and Mining Soc. of South Africa*, 52 (6): 119–139.

Matheron, Georges. 1962. Traité de géostatistique appliquée, Editions Technip.

Zhang R. 2005. Spatial Variability Theory and Its Application. *Science Press,* Beijing.

Isaaks E. H., Srivastava R. M. 1989. An introduction to applied geostatistics, *Oxford University Press*, Inc.

Webster R., Oliver M. A. 2007. Geostatistics for environmental scientists, 2nd ed. John Wiley & Sons, Ltd.

Pyrcz, M.J., Deutsch, C.V. 2003. The whole story on the hole effect. In: Searston, S. (Ed.), Geostatistical Association of Australasia. Newsletter 18, West Perth, WA, Australia, 18 pp.